Supplementary Material

M1. Text Vectorization

In this study, we used Term Frequency-Inverse Document Frequency (TF-IDF) and Bag of Words (BoW) extraction methods to vectorize text data. TF-IDF measures the importance of a word in the document as the product of Term Frequency (TF) of the term in that document and the Inverse Document Frequency (IDF) of the term throughout the corpus, as shown in the equation below.

$$TF - IDF = \ tf_d^i * \log(\frac{N}{df^i})$$

Where, tf_d^i is the number of times the term i occurs in the document d, N denotes the number of documents in the corpus, df^i is the number of documents with the term i.

BoW describes documents by the frequency of words. BoW representation of a document is used for probabilistic topic models, such as Latent Dirichlet Allocation (LDA) models, because LDA estimates probability distributions for words in topics and topics in documents, which does not necessarily require the weighting of TF-IDF text representation (Blei et al., 2003).

M2. Topic Modeling

We used Latent Dirichlet Allocation (LDA) to explore topics discussed in the COVID-related tweets. The LDA model defines a collection of D documents (i.e., tweets in this study) as a corpus. A document is a sequence of N tokens/words. The model assumes the following generative process of a corpus with D documents and K topics (Blei et al., 2003; Hoffman et al., 2013).

1. For each topic, $k \in K$, draw a probability distribution of a word/token appearing in topic k as β_k that follows Dirichlet distribution with parameter η (i.e., the prior distribution over the words, input by the model developer).

2. For each document $d \in D$, draw a probability distribution of topic proportions that the document d belongs to as θ_d that follows a predefined Dirichlet distribution with parameter α (i.e., the prior distribution of topic proportions, input by the model developer).
3. For each word/token i in document d:
    a. Draw the topic assignment z_{di} that follows the Multinomial distribution of θ_d
    b. Draw the observed word/token w_{ij} that follows the Multinomial distribution of $\beta_{z_{di}}$

The posterior distributions of document-topic distribution θ and topic-word distribution β can be approximated using various algorithms, such as the online variational Bayes algorithm used by the Scikit-learn 1.0 package (Blei et al., 2003; Hoffman et al., 2013; Pedregosa et al., 2011). The selection of parameters for prior distributions η and α will influence the model outputs. In this study, we used the default parameters in the Scikit-learn package for model estimation, i.e., both η and α are initialized as the $1/k$.

M3. Cosine Distance Measure

We captured the transit agencies' tweeting repetitiveness using the text cosine distance measure, which is the L2-normalized dot product of text vectors (Schütze et al., 2008). Let x and y denote two vectorized documents. Their cosine distance c is calculated using the following equation:

$$c(x, y) = \frac{xy^T}{||x||\,||y||}$$

Cosine distance measure ranges from 0 to 1. The larger the measure, the more similar set of words are used by the two documents/tweets. The cosine distance measure has been used to examine similarity-based context classification (Fu et al., 2015; Paule et al., 2019) and clustering and content recommendation with similar words (Tajbakhsh & Bagherzadeh, 2016).

M4. Streamgraph visualization

The streamgraph in Figure 3 (left) demonstrates the frequency of the six topics (each topic is a different color) and the total volume of COVID-related Tweets. This chart applies a Gaussian smooth kernel filter to the original data (i.e., number of COVID-related tweets by topics) to smooth the trend, which is considered more appropriate to visualize in time series data (Byron & Wattenberg, 2008). The total number of COVID-related tweets by topic is approximated by the upper and lower points of the graph at any point in time.

Figure S1: Number of twitter handles by transit agencies (transit agencies ranked by VOMs in descending order)

Table S1: Examined hyperparameters by machine learning model

Model	hyperparameters (names from the scikit-learn 1.0 package)	Tested list/range
Random Forest Classifier	n_estimators	range(10,250,10)
	max_features	range(5,30)
	max_depth	range(2,10)
	min_samples_split	range(2,30)
	min_samples_leaf	range(2,20)
Gradient Boosting Classifier	n_estimators	range(10,250,10)
	max_features	range(5,30)
	max_depth	range(2,10)
	min_samples_split	range(2,30)
	min_samples_leaf	range(2,20)
	learning_rate	range(0.0001, 1)
Support Vector Machine	kernel	[‘linear’, ‘poly’, ‘rbf’, ‘sigmoid’]
	degree	[2,3]
	C	range(0.01,1)
Logistic Regression	penalty	[‘l1’, ‘l2’, ‘elasticnet’, ‘none’]

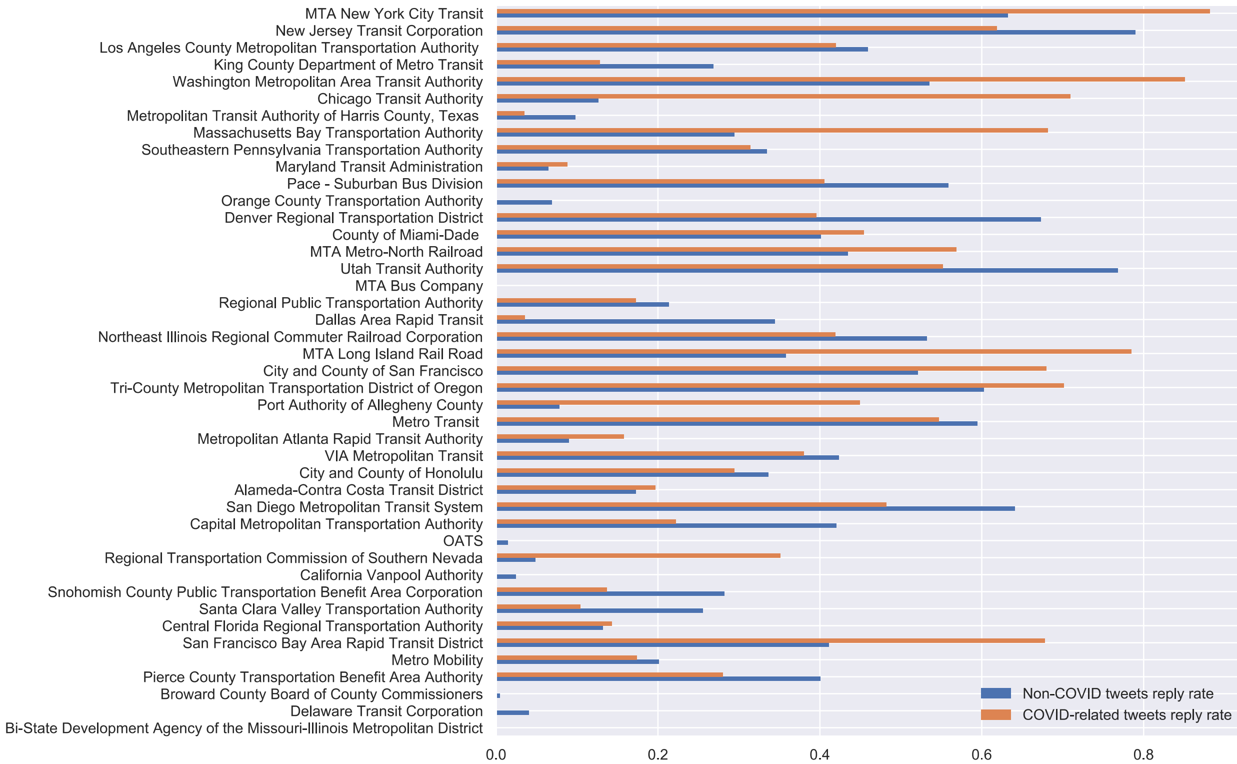

Figure S2: **GENERAL ACTIVITY:** Reply tweets rate for non-COVID-related and COVID-related tweets

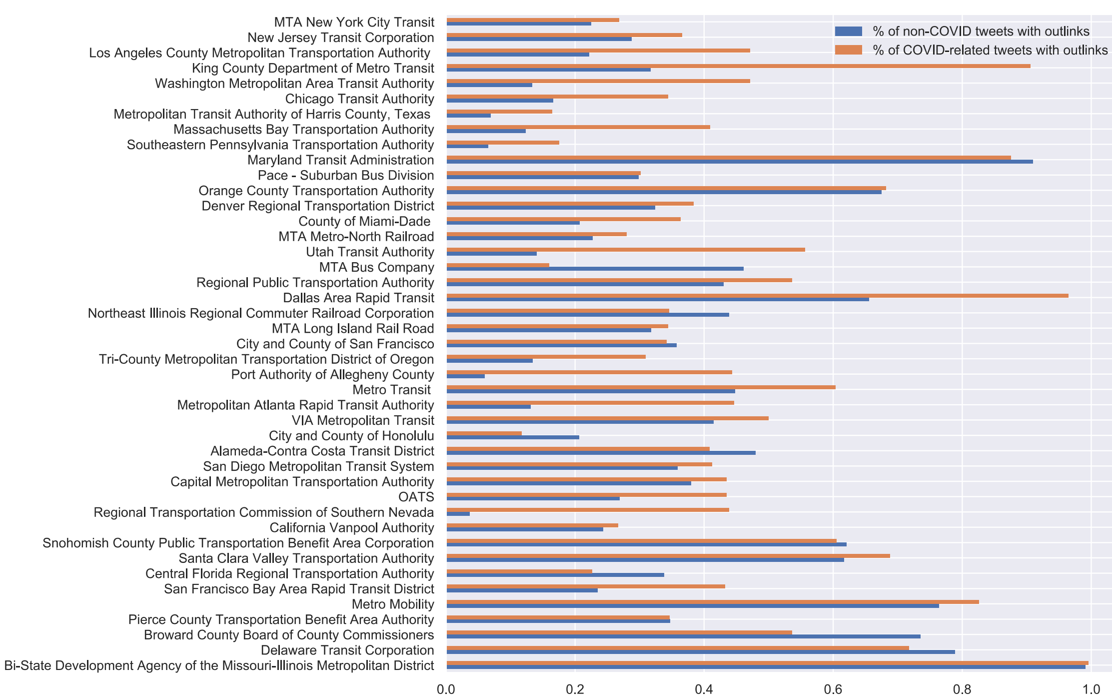

Figure S3: **GENERAL ACTIVITY:** Percent of tweets using outlinks for non-COVID-related and COVID-related tweets

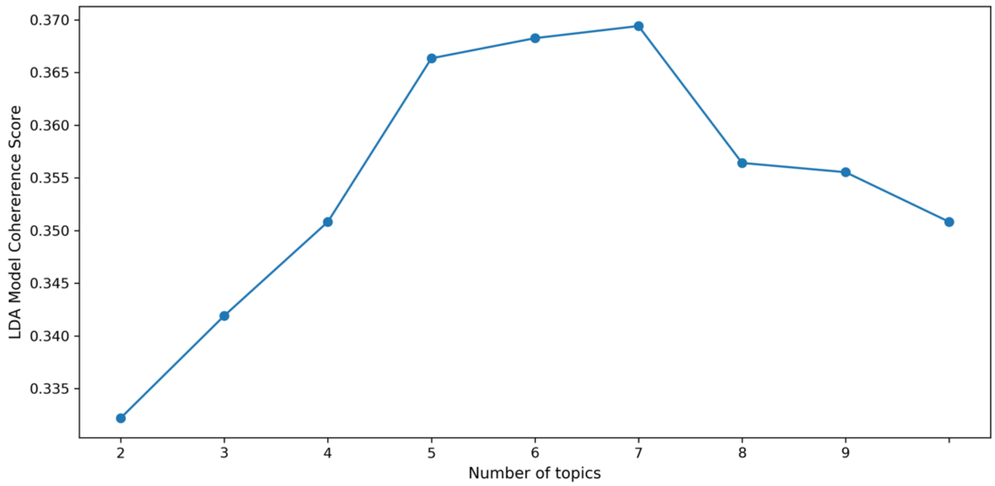

Figure S4: LDA model coherence score by number of topics

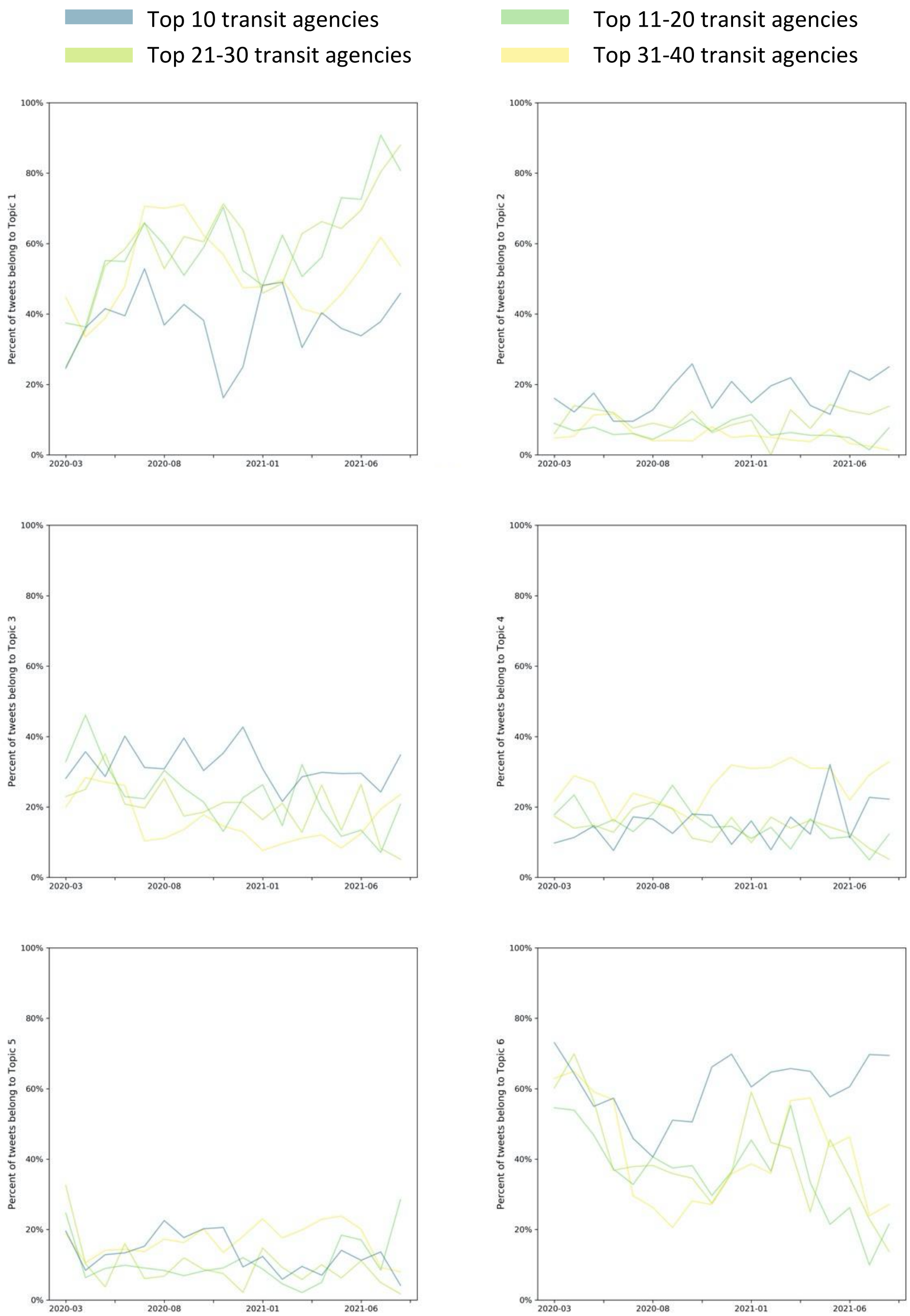

Figure S5: **CONTENT ANALYSIS:** The percent of monthly tweets belonging to each topic by transit agencies during the pandemic (line color corresponding to the VOMs volume).

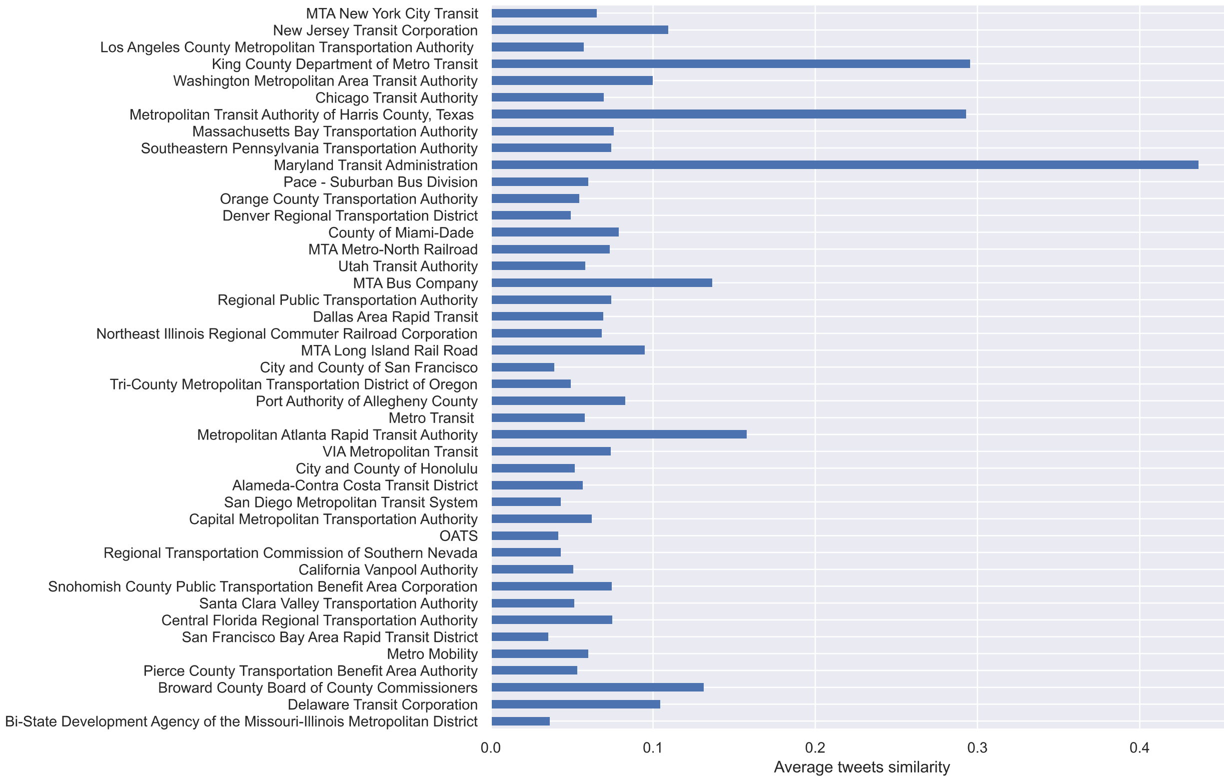

Figure S6: **CONTENT ANALYSIS:** Average tweets similarity by transit agencies

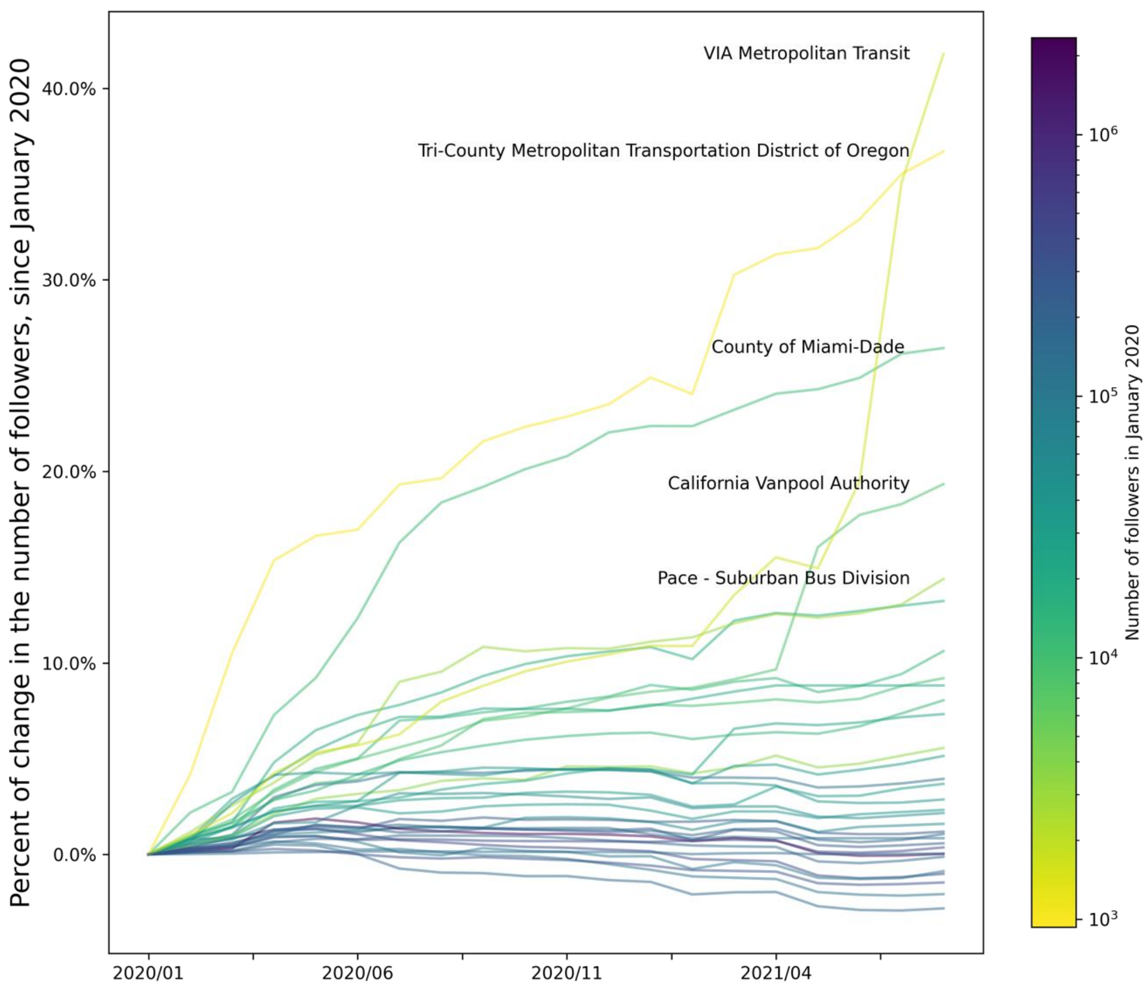

Figure S7: **RIDER INTERACTION:** Follower changes during the pandemic by transit agency. Line color corresponding to Number of Followers in January 2020. The darker the more follower the agency has (which is highly correlated with the transit agency size)

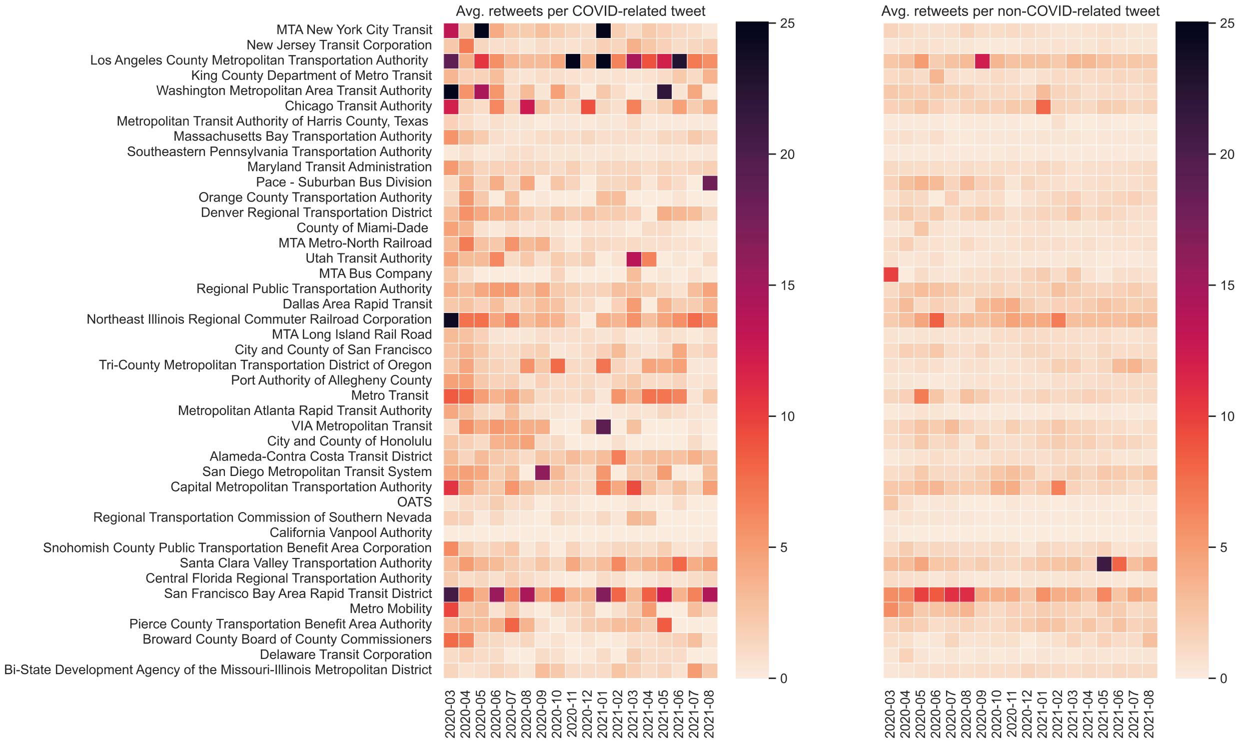

Figure S8: **RIDER INTERACTION:** Number of retweets per COVID-related (left) and non-COVID-related tweets (right)

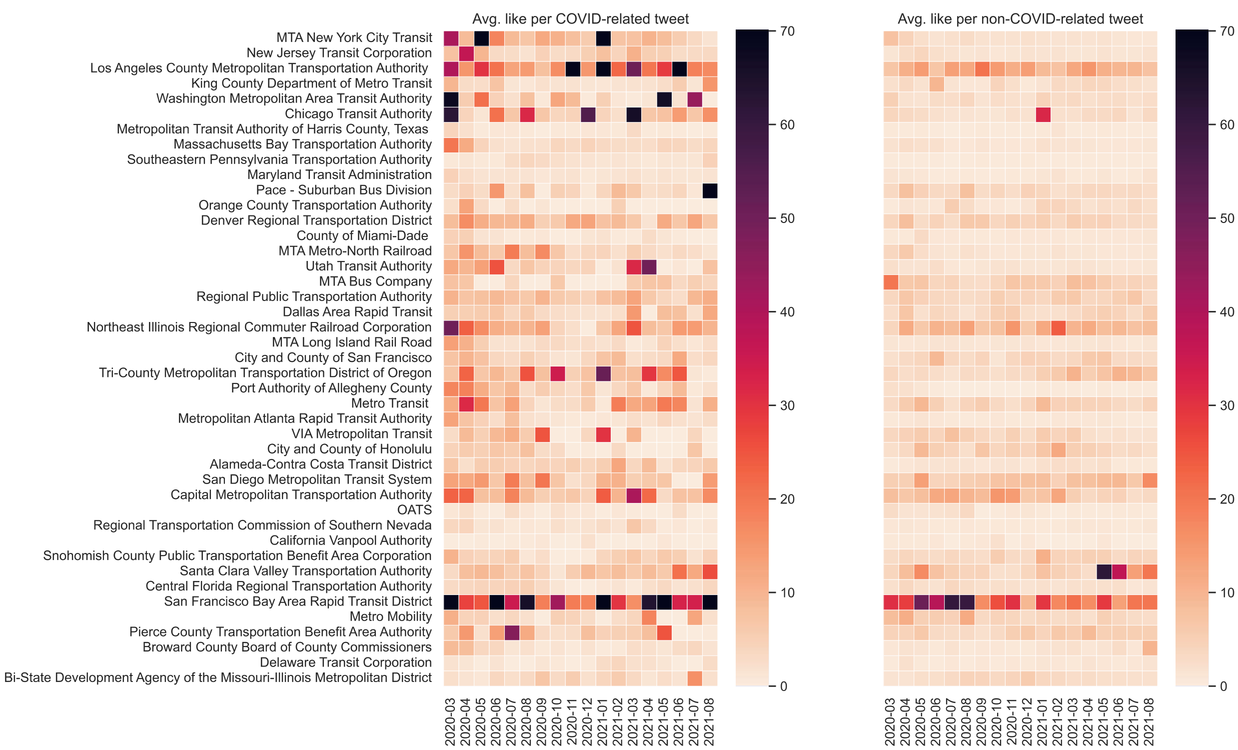

Figure S9: **RIDER INTERACTION:** Average number of likes per COVID-related and non-COVID-related Tweets

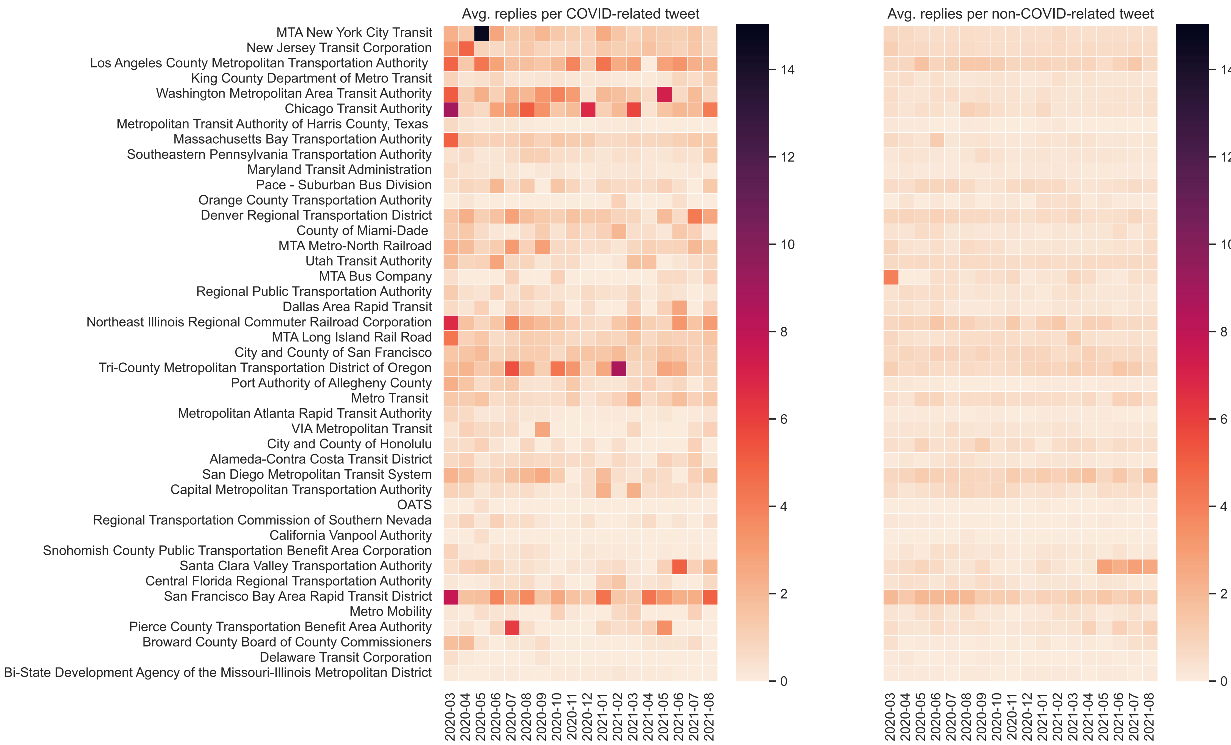

Figure S10: **RIDER INTERACTION:** Average number of replies per COVID-related and non-COVID-related Tweets

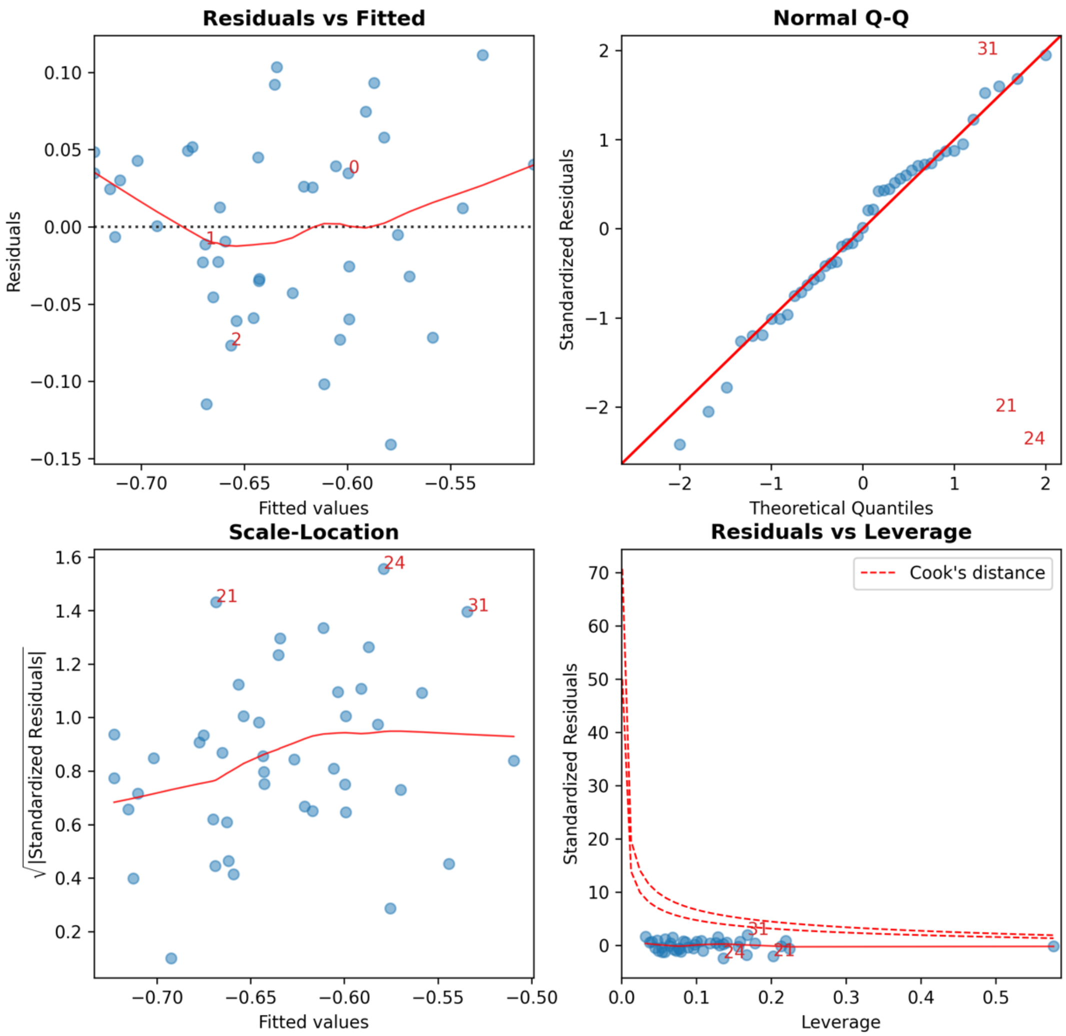

Figure S11: Percentage of changes in the number of followers model diagnostic charts.

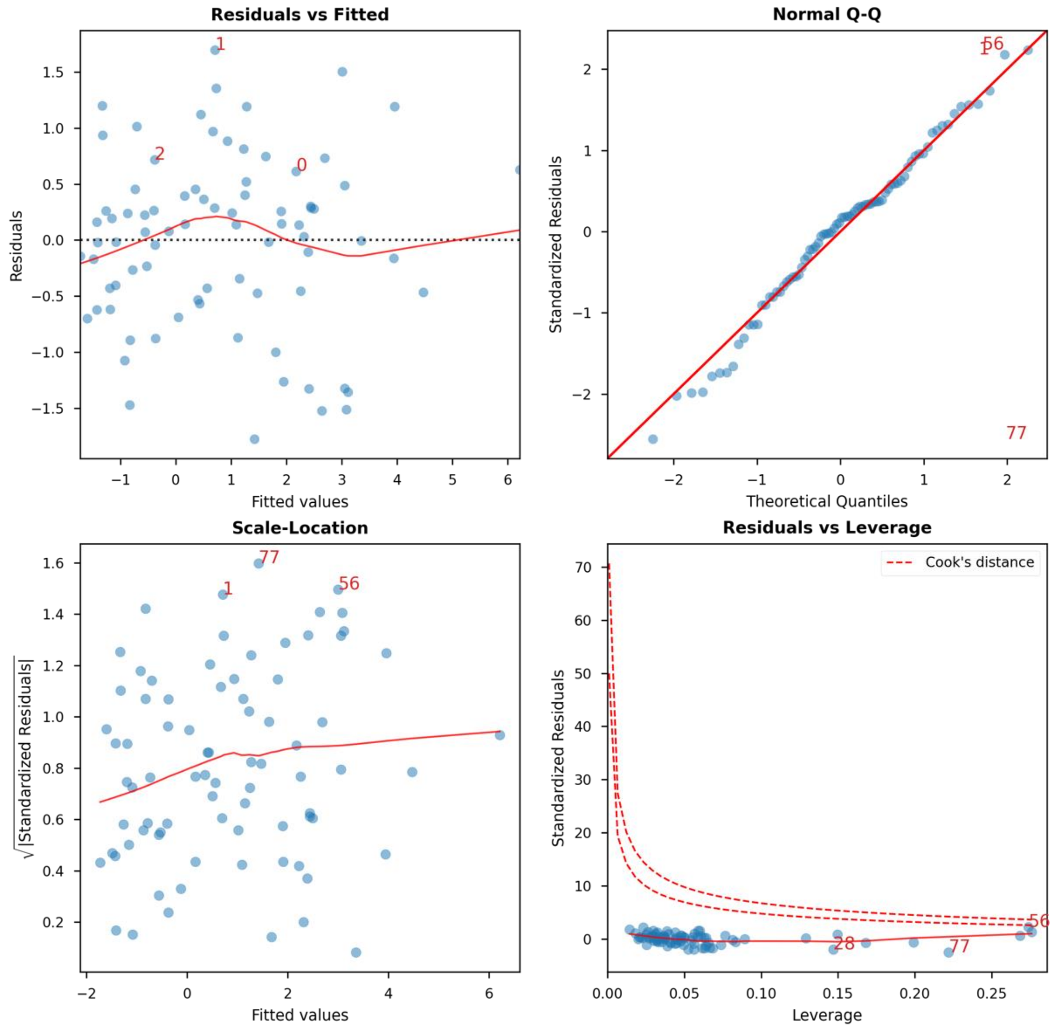

Figure S12: Average likes counts per COVID-related model diagnostic charts.

References

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, *3*, 993–1022.

Byron, L., & Wattenberg, M. (2008). Stacked graphs–geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics*, *14*(6), 1245–1252.

Hoffman, M. D., Blei, D. M., Wang, C., & Paisley, J. (2013). Stochastic variational inference. *Journal of Machine Learning Research*, *14*(5).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.